

This is a preprint of the following article, which will be available from <http://octadlab.ust.hk>.

D. Kim, A. Seth, and R. P. Liem. Data-enhanced dynamic flight simulations for flight performance analysis. *Aerospace Science and Technology*, 121:107357, 2022. doi:10.1016/j.ast.2022.107357

The published article may differ from this preprint, and is available at the DOI above.

Data-enhanced dynamic flight simulations for flight performance analysis

Dajung Kim* Arjit Seth† Rhea P. Liem‡

The Hong Kong University of Science and Technology (HKUST), Hong Kong SAR

Abstract More often than not, airlines use aircraft in operating scenarios beyond their optimal design conditions, which negatively affects their performance characteristics. These effects are statistically reflected in flight operational data, which are indirectly constrained by air transportation management as well as aircraft design. In this research, we develop a data-enhanced methodology for modeling dynamic flight simulations of aircraft to enable accurate estimation of performance parameters. The relevant flight phases and constraints of these simulations are determined by employing supervised machine learning on the flight data. The methodology is demonstrated by simulating flights across representative short-, medium-, and long-haul sectors using data shared by our airline partner. We compare the fuel burn and flight time calculation results with those from the Bréguet range and endurance equations, from an available open-source flight performance model, and from the reference data for validation. The developed dynamic flight simulation model is designed in such a way to enable accurate flight performance analysis even when high-fidelity force analyses and control models are absent, which is common in preliminary design frameworks. This will further enable incorporating flight data into aircraft design processes.

Nomenclature

(L/D)	Lift-to-drag ratio	τ	Throttle factor
α	Angle of attack	$\vec{\omega}$	Angular velocity
β	Speed of sound	\vec{a}	Acceleration
$\Delta()$	Change in quantity	\vec{C}_A	Aerodynamic coefficients
η	High-lift device configuration	\vec{F}_A	Aerodynamic forces
$\mathcal{R}(\vec{y}; \vec{x})$	Nonlinear residual equations with inputs \vec{x} and outputs \vec{y}	\vec{F}_T	Propulsive forces
ψ, θ, ϕ	Euler angles	\vec{g}	Acceleration due to gravity on Earth
ρ	Density of air	\vec{r}	Coordinates of the aircraft
σ	Air density ratio	\vec{V}	Velocity
		BPR	Engine bypass ratio

*Ph.D. Candidate, Department of Mechanical and Aerospace Engineering.

†Research Assistant (M.Phil.), Department of Mechanical and Aerospace Engineering.

‡Assistant Professor, Department of Mechanical and Aerospace Engineering.

C_D	Drag coefficient	S	Wing area
C_L	Lift coefficient	t	Time
c_T	Specific fuel consumption	T_{amb}	Ambient temperature
h	Altitude	V_{stall}	Stall speed
M	Mach number	V_{CAS}	Calibrated airspeed
m	Aircraft mass	V_{TAS}	True airspeed
p_{amb}	Ambient pressure	W	Aircraft weight
R	Flight range		

1 Introduction

Airlines tend to purchase oversized aircraft and use them under diverse operating conditions in favor of fleet commonality, notwithstanding the possibility of higher fuel consumption and operating costs [1]. Recently, the need for flexibility has increased due to the changes in the air transportation landscape brought about by the COVID-19 pandemic [2]. The pandemic has significantly affected the aviation industry resulting in approximately 350 billion USD revenue loss per year [3], and thus airlines are undergoing many changes in the search for sustainability. One of them is the decrease in long-haul international flights of long-range large aircraft [4], while short-haul flights remain popular. Passenger aircraft have also been refitted into cargo aircraft [5]. Considering these changes, it is becoming increasingly common to operate an aircraft outside its on-design operating points and with sub-optimal takeoff weights [4]. Moreover, previous predictions show that many aircraft currently in service worldwide are close to the end of their commercial lives, and most major airlines need to replace their several hundred heritage aircraft in the near future [6]. In the context of such rapid changes in the aviation industry, it is becoming more crucial to estimate aircraft performance at the earlier stages of the aircraft design process under varying flight conditions to reduce extra costs. To accurately estimate aircraft performance, detailed aircraft information is required. This information, however, is typically not readily available. To address this issue, we present a data-enhanced flight simulation model that can estimate performance factors. By utilizing flight data in the model construction, we can reduce the amount of required aircraft information in the model derivation. The model aims to obtain high accuracy, especially in measurement of fuel consumption and flight time. The effectiveness of this model is evaluated by studying the performance of a long-range large aircraft for short-, medium-, and long-haul flights with varying payloads.

Important aircraft performance factors include takeoff and landing distances, rate of climb, ceiling, speed, payload, range, duration, and fuel economy [7]. Based on our observations, models for predicting these flight performance factors can be broadly classified into three types: (1) data-driven models, (2) physics-based models, and (3) hybrid models. Their characteristics and limitations will be briefly described below. Data-driven flight-performance models rely on a vast amount of historical flight data, from which insightful trends and patterns are discovered. They are commonly used to estimate and analyze fuel economy and flight time, among other performance factors. Fuel economy is defined by how much fuel is consumed when flying in a given payload and range configuration and is one of the most important factors from the perspective of airlines' finances [8] along with flight time, which is important for flight scheduling. However, most data-driven fuel consumption models use other performance factors as inputs, such as flight time [9, 10, 11]. Data-driven flight time models mostly focus on expected time of arrival [12, 13], the correlated delay [14,

15, 16], and overall traffic flow [17, 18] rather than aircraft performance factors such as endurance. Hence a data-driven model is not suitable for simultaneously estimating fuel consumption, flight time, and other flight performance factors.

In the governing equations of aircraft dynamics, physics-based models apply numerous simplifying assumptions to make analyses computationally tractable, such as first modeling the aircraft as a rigid body [19]. This assumption enables detailed aircraft performance analyses including stability analysis and fuel burn over time [20], but the model generally requires control settings; it is widely used in flight simulators for pilot training [21]. The second is a point-mass assumption, which is useful when the focus of the study is on mass transport in the form of a trajectory. It is mainly used for comprehensive aircraft performance analysis in air transportation studies such as the Open Aircraft Performance model (OpenAP) [22]. The third is a steady, level flight assumption, which leads to the classic Bréguet range equation [23]. This assumption is commonly used for mass and endurance estimation at preliminary aircraft design stages for a single design point in software packages such as Multidisciplinary Airplane Research Integrated Library (MARILib) by ENAC, IRT Saint-Exupéry, and Airbus [24]. A limitation of this last approach is that it neglects important accelerated segments of flight which contribute to fuel burn in performance analysis.

We classify hybrid models into two kinds: one in which data models are constructed or trained using results from physics-based models, and one in which physics-based models use data in their formulations. Recent research by Yu *et al.* presented a flight dynamics model built with a deep residual recurrent neural network model using data obtained from solving the governing equations of motion under a rigid-body assumption [25]. Seymour *et al.* proposed a model combining a point-mass, rigid-body, two-dimensional flight dynamics model and a data-based reduced-order fuel-burn approximation model [26]. Their models had the benefit of low computational time but could not overcome the limitations imposed by the assumptions of the physics model. On the other hand, a data-enhanced physics model uses data to correct errors from the assumptions of the physics-based model. Such models based on the Bréguet range equation mostly focused on correcting errors in flight segments other than cruise. Randle *et al.* modified the Bréguet range equation by applying range, lost-fuel, and recovery-fuel factors derived from flight data [27]. Yanto and Liem recently developed a fuel burn estimation model which used the Bréguet range equation for cruise combined with data-corrected models for the climb and descent segments [10]. These methods can predict fuel burn with fast computation time and high accuracy. However, they cannot be applied to new aircraft types as they cannot consider changes in the physical parameters of the aircraft without its flight records. A data-enhanced physics model based on the flight equations of motions can compensate for this drawback, especially considering dynamic stability constraints in the context of multidisciplinary design optimization [28, 29]. Chatterji used the flight equations of motions with a point-mass assumption, and a fuel-flow model derived from flight data to estimate fuel consumption [30]. This method estimated the takeoff weight within 1% error and included design of the altitude controller, which was a complex requirement. Lyu and Liem solved two-dimensional kinetic equations transformed into range equations for each flight segment, with data-driven mission parametrization [31]. The segmentation procedure for the mission parametrization focused on cruise and step climbs with data analysis for accurate fuel-burn prediction, but heuristic rules were used to determine other flight segments. The formulation conveniently did not require a control model, but also did not guarantee compatibility of state variables between boundary conditions of the segments, only for the weights. In our present study, we address the aforementioned limitations by adopting machine learning into the segmentation procedure and developing a methodology which incorporates data processing into the formulation of constraints and boundary conditions in physics-based models.

In this paper, we present a new formulation for a data-enhanced flight simulation physics model

via solution of kinetic and kinematic differential equations of motion under a point-mass assumption. The relevant constraints and boundary conditions are formulated based on flight data provided by our airline partner, Cathay Pacific Airways Limited. The model aims to address limitations in purely data-driven models and physics-based models in the context of flight performance analysis. By applying supervised machine learning algorithms on these data to characterize the mission profiles, we can derive an accurate flight performance model by relying only on minimal information from the aircraft, such as elementary engine parameters and the wing’s geometric characteristics. The formulation constrains the flight trajectory by the averaged vertical speed, target speed, or altitude points for specific segments obtained from analyzing flight data using the XGBoost classification model. These constraints and boundary conditions are imposed on the equations of motion to make them well-posed, and the solutions contribute to high accuracy in predicting fuel consumption and flight time even without wind corrections, high-fidelity force analyses, or control models. Using this approach, the dynamic flight simulations of our model can mimic aircraft operations with a control unit following waypoints, which is observed to improve the accuracy of performance estimation compared to current methodologies used in preliminary design. Hence, this methodology can also be used for preliminary design of future aircraft to suit current market demands of air travel, and the formulation can also be used to reduce the number of unknowns in trajectory optimization studies of aircraft.

This paper is structured as follows: Section 2 covers the data and physics modeling, and how these are integrated into a data-enhanced flight simulation methodology. Section 3 presents the results of the methodology for flight simulations conducted over various routes for a single aircraft. It also discusses performance analyses using the results while comparing them to other methodologies with validation. Section 4 concludes with a summary of the formulation and results.

2 Methodology

In this study, we incorporate flight data to enhance our physics model, which solves the governing equations of flight dynamics. The data-set consists of detailed flight information acquired from Quick Access Recorder (QAR) data provided by our airline partner. The flight point data are recorded every 10 seconds, which are used to train and validate a flight segmentation model. We use a machine learning classification model to analyze and classify raw flight data into different segments based on the values of and the corresponding changes in speeds, positions and other state variables. The flight point data and their corresponding classified segments are analyzed to determine the speed and altitude at the start- and end-points of each segment and the representative values of vertical speed. The physics model performs flight simulations using information gleaned from the segmentation and the aforementioned data analysis. Performance factors such as mass, flight time, and other state variables can hence be obtained at different points along the flight trajectory from the simulations. The overview of the framework outlining this process is provided in Fig. 1.

In Section 2.1, the method to construct the data-based segmentation model is presented with the corresponding result validation. In Section 2.2, the physics model is described in detail. Section 2.3 explains the process of analyzing flight data and how they are used as constraints in the physics model.

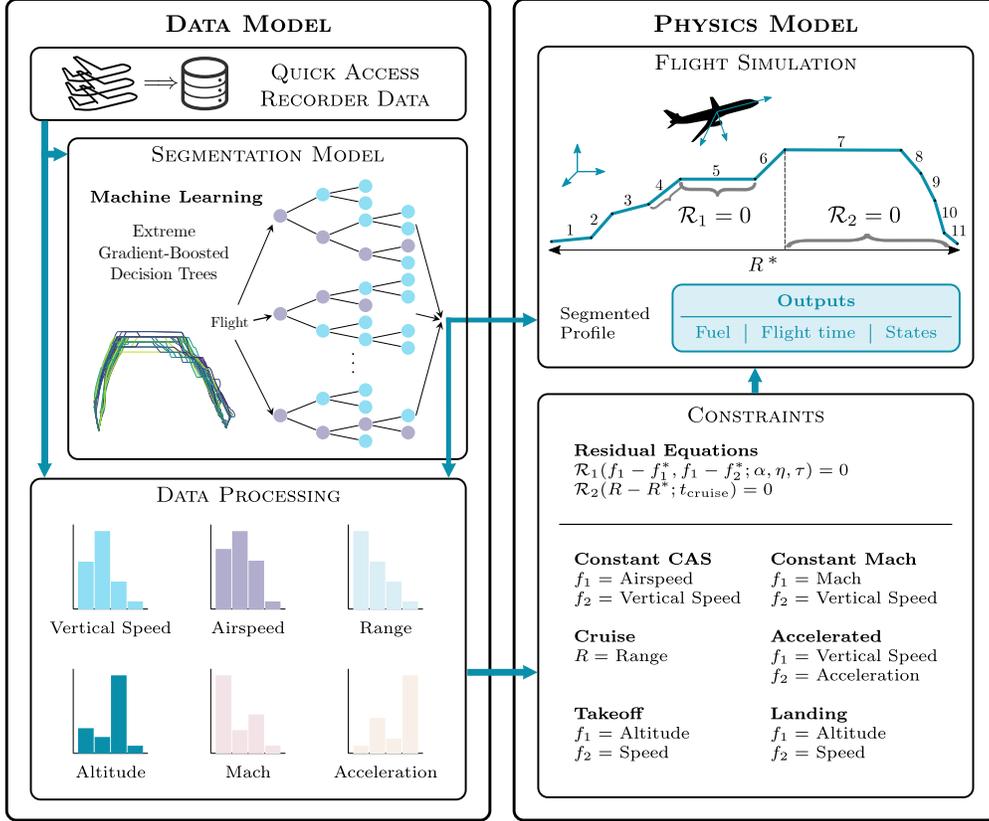


Figure 1: Overview of the proposed methodology for data-enhanced flight simulations.

2.1 Mission specifications

The mission specifications refer to the flight scenarios, which are inputs to the flight simulation procedure. To define a complete flight scenario, we need information on the total mission range, the combination of flight segments separated by characteristics of altitude and velocity variation, and the target altitude and velocity at the beginning and end of each flight segment. Fig. 2 illustrates the generic mission specifications for a flight. The mission range is determined by the ground distance between the origin and destination. Nine flight segments are considered: takeoff, accelerated climb, constant calibrated airspeed (CAS) climb and descent, constant Mach number climb, descent and cruise, decelerated descent, and landing.

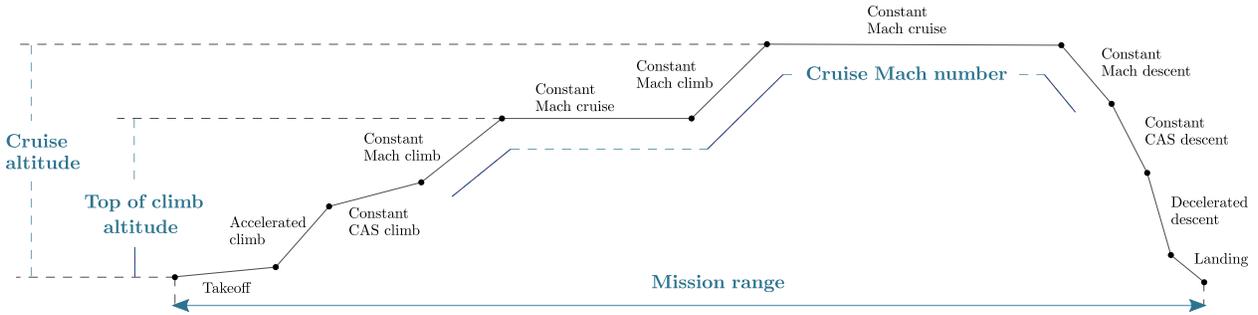


Figure 2: A mission profile example showing all flight segments (not to scale).

2.1.1 Classification model generation

In this work, we use a classification model based on eXtreme Gradient-Boosted decision trees (XGBoost) [32] as the supervised learning model to cluster flight point data into segments. The XGBoost classifier increases the accuracy of the model by generating gradient-boosted decision trees with regularization techniques. Along with the bagging method, which constructs decision trees with randomly selected subsets from the training data and aggregates them later, a weighted quantile sketch and an approximate greedy algorithm are applied to generate an accurate model with fast training time, even for large data-sets. The automatic feature selection applied in XGBoost guarantees high prediction accuracy to a certain extent even if the sensitivities of the results to the selected feature inputs are low.

A subset of the flight point data obtained from the airline’s flight data-set is used to train this clustering algorithm. The classification model receives variations of state variables, which are throttle factor τ , vertical speed V_z , Mach number M , altitude h , calibrated airspeed V_{CAS} , and true airspeed (TAS, V_{TAS}) as inputs, and returns the segment types as outputs. A classification model is potentially sensitive to flight conditions under small variations, such as gusts or collision avoidance maneuvers, and may classify a perceptibly single flight segment as multiple small segments. This can be interpreted as model over-fitting, and the regularization techniques employed in the XGboost classification model help reduce such sensitivities. Fig. 3 describes the constructed nonlinear data-based model using Shapley additive explanations pertaining to these input factors [33]. The size of each bar in Fig. 3 indicates the sensitivities of the prediction output to the input factors. The lengths of subdivisions of the bars indicate the impact of the respective variable on the flight segment in the corresponding color. Overall, the throttle factor of the model has the greatest influence on segment classification, as can be seen in the total length of its corresponding bar. The throttle factor is also observed to be the dominant factor in determining constant CAS descent, decelerated descent, constant Mach climb, and landing segments by comparing the lengths of the color-specific bars.

2.1.2 Classification model verification

The segmentation model is trained with 120 181 flight point data and verified with 23 194 non-overlapping points. The prediction error is minimized by tuning five hyperparameters (i.e. number of trees, maximum depth, minimum sum of weights, subsample ratios, and step-size shrinkage) as listed in Table 1. The first column represents the number of gradient-boosted decision trees. The second represents the maximum depth of the nodes in a single decision tree. The third represents the minimum sum of weights, which corresponds to the minimum number of instances required in each node. The fourth represents subsample ratios, which indicate the ratio of the cardinalities between a selected subset to the training data-set used to grow trees. The last represents the step-size shrinkage, which is a regularization parameter used in the update to prevent over-fitting. The values of the hyperparameters which are used in building the model are denoted distinctively in the table.

Fig. 4 shows the verification results of the model for each segment, in terms of precision, completeness, and weighted harmonic mean, using the classification report visualizer from the Python package `yellowbrick`. This tool is specifically designed to assist in evaluating the performance of machine learning models, especially in terms of stability and predictive value [34]. *Precision* is a measure that quantifies the ratio of true positives to the total number of true positives and false positives corresponding to a particular classified segment. Using the takeoff segment as an example, the precision corresponds to the correct predictions of takeoff segments against the total number of times the model classifies a segment as a takeoff segment. *Completeness* is the ratio of true

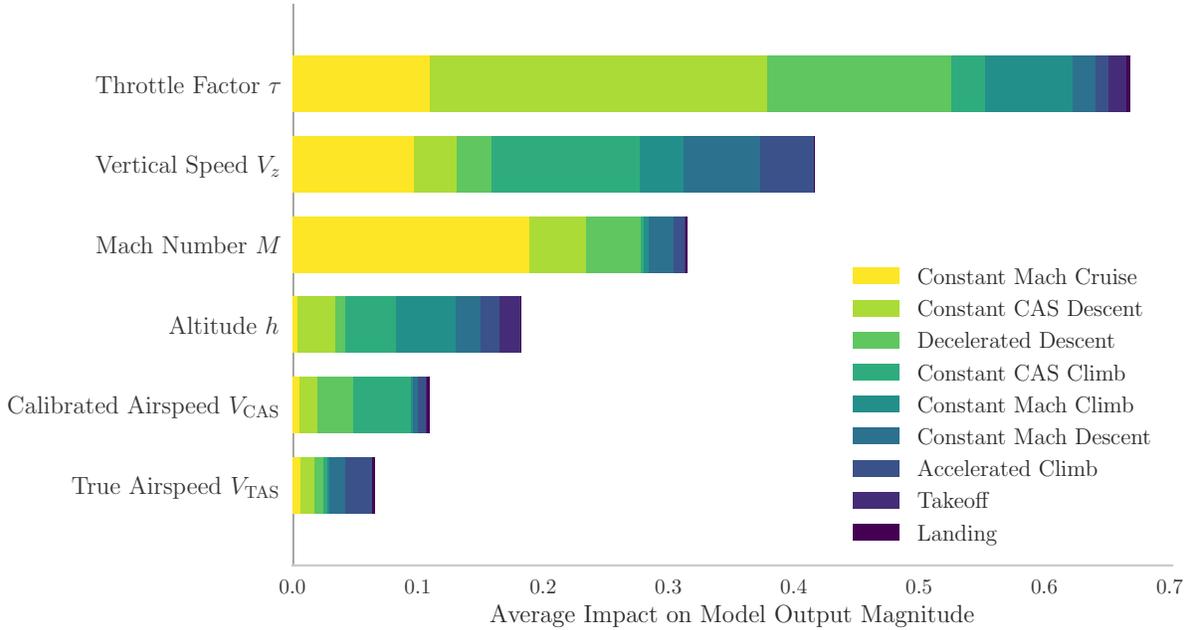


Figure 3: Results of average impact of the state variables on the classification model.

Table 1: The combinations of tested hyperparameters.

Number of trees	Maximum depth	Minimum sum of weights	Subsample ratios	Step-size shrinkage
200	9	5	7/10	0.3
400	10	6	8/10	0.2
600	11	7	9/10	0.1
800			10/10	0.05
1 000				0.01
				0.005

positives to the sum of true positives and false negatives. In the same example, the completeness corresponds to the takeoff segments predicted correctly against the actual takeoff segments. These two measures are combined by calculating their *weighted harmonic mean*, which is shown in the last column of Fig. 4. The values for these three measures range from zero to one, as indicated in the color bar. The results presented in Fig. 4 show that the segmentation model identifies different segments with a high level of accuracy. Although the segment-by-segment accuracy varies with a minimum weighted harmonic mean of 86%, the segmentation model achieves a global accuracy level of 96%, which is deemed sufficient for the purpose of this study.

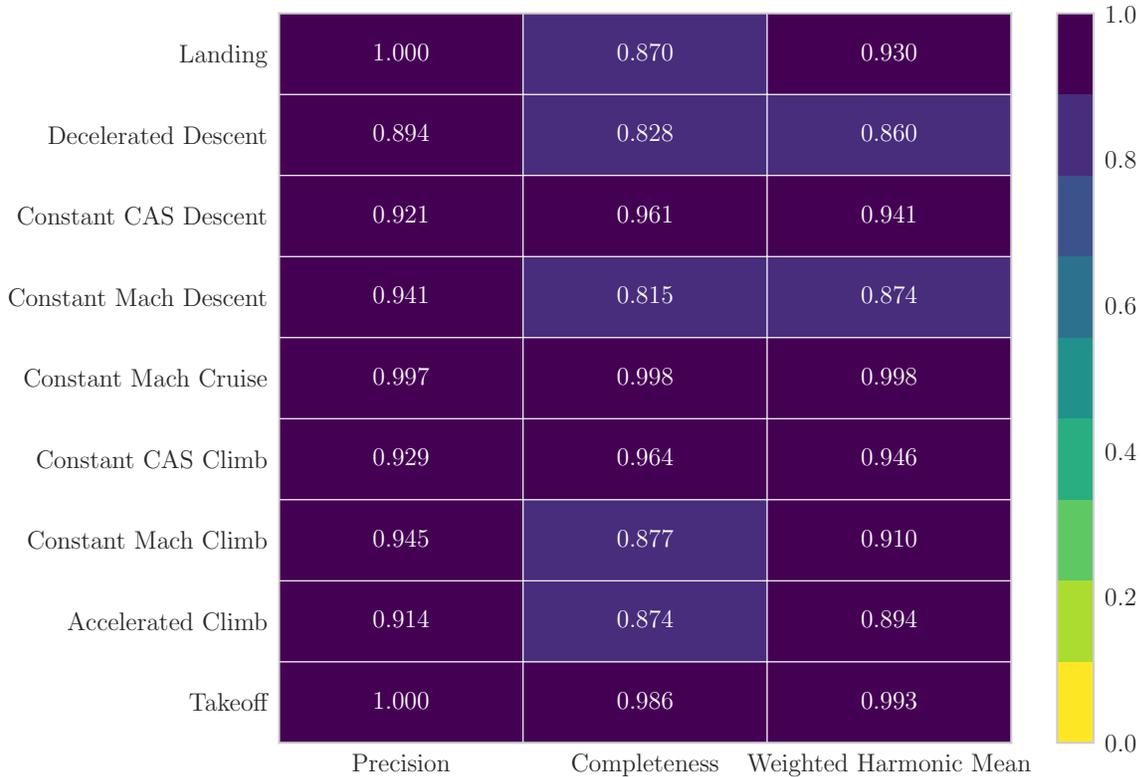


Figure 4: Results of the XGBoost classification analysis.

2.2 Flight simulation model

In the current approach, the governing differential equations of motion for an aircraft are discretized and integrated in time as an initial-value problem. The relevant components analyze the flight dynamics, gravitation, atmosphere, aerodynamics, propulsion, and mass models. Each of these components will be described in detail in Sections 2.2.1 to 2.2.5. The process is depicted in Fig. 5 using the eXtended Design Structure Matrix (XDSM) representation developed by Lambe and Martins [35]. The external inputs, such as the angle of attack α , the throttle factor τ , the high-lift device configuration η , are determined via optimization procedures which solve residual equations to satisfy the equations of motion under constraints imposed by the data model, discussed in Section 2.3.

2.2.1 Flight dynamics

The translational equation of motion governing the flight dynamics is presented in Eq. (1), where the aircraft is modeled as a point-mass rigid body.

$$\vec{F}_T + \vec{F}_A + m\vec{g} = m(\vec{a} + \vec{\omega} \times \vec{V}) \quad (1)$$

In this formulation, mixed boundary conditions are imposed on the differential equation for different segments, which will be further described in Section 2.3. The acceleration \vec{a} is determined by the mass m , propulsive and aerodynamic force components \vec{F}_T , \vec{F}_A , gravitational acceleration \vec{g} ,

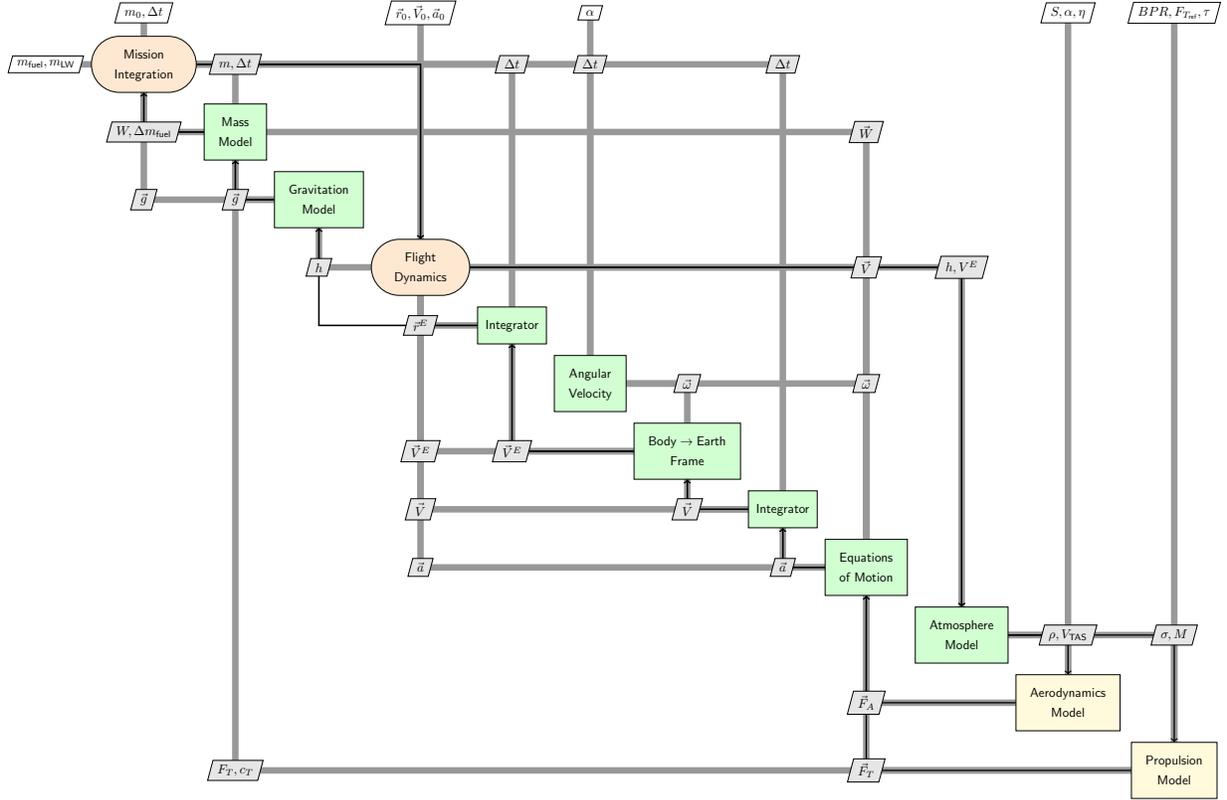


Figure 5: XDSM of proposed flight simulation procedure with models.

velocity \vec{V} and angular velocity ω obtained as explicit computations, hence the changes in velocity $\Delta\vec{V}$ and position $\Delta\vec{r}$ of the aircraft in the body frame can be calculated via integration.

Eq. (1) is discretized in time using finite differences, and numerical integration is performed using the forward Euler method, in which the number of elements in the integration differ depending on the type of segment. The discretization is represented as t_k for $k = 0, \dots, K$, where $k \in \mathbb{N}$. Hence for N flight segments, the total number of discretized points for one flight simulation is given by $K = \sum_{p=1}^N M_p$, where M_p represents the number of discretized points in the p th flight segment. A functional which is dependent entirely on dynamical variables at a given time-step, e.g. $f(x(t_k), y(t_k))$, is notated as a function of the time-step itself $f(t_k)$, and in some equations its value is denoted as f_k . The numerical approximation for integration is denoted in Eq. (2), with $-i$ subscripts for initial values and $-f$ for final values. The angular velocity is calculated by computing the differences in specified angles between consecutive time-steps, instead of solving the rotational equation of motion as is usually done in a flight simulation analysis. This formulation is adopted to utilize the data model presented in Section 2.1 and eliminate the requirement of a control model. The method for determination of the angles and the formulation are described in Section 2.3.

$$\begin{aligned} \Delta\vec{V} &= \vec{V}_f - \vec{V}_i = \int_{t_{prev}}^{t_{next}} \vec{a}(t) dt \approx \sum_{n=0}^N \vec{a}(t_n) \Delta t \\ \Delta\vec{r}^E &= \vec{r}_f^E - \vec{r}_i^E = \int_{t_{prev}}^{t_{next}} \Delta\vec{V}^E(t) dt \approx \sum_{n=0}^N \Delta\vec{V}^E(t_n) \Delta t \end{aligned} \quad (2)$$

The linear and angular velocities expressed in the body frame are converted into their representations in the Earth frame, with $-^E$ superscripts, using the Euler angle coordinate transformation shown in Eq. (3) at the given instant of time. The rotation transformation matrices for Euler angles ψ , θ , ϕ (corresponding to yaw, pitch and roll angles, respectively) are presented in Eq. (4). The transformed velocity is used as an input to the aerodynamics and propulsion models to determine the necessary external forces. This transformation also provides the global position of the aircraft, which is used as an input into the atmosphere, gravitation, and propulsion models subsequently presented.

$$\begin{aligned}\Delta\vec{V}^E(t_n) &= R_{\psi\theta\phi}(t_n)\Delta\vec{V}(t_n) \\ \vec{\omega}^E(t_n) &= R_{\psi\theta\phi}(t_n)\vec{\omega}(t_n)\end{aligned}\tag{3}$$

$$R_{\psi\theta\phi} = \begin{bmatrix} \cos\psi & -\sin\psi & 0 \\ \sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & -\sin\phi \\ 0 & \sin\phi & \cos\phi \end{bmatrix}\tag{4}$$

2.2.2 Atmosphere and gravitation models

The inputs to the aerodynamics and propulsion models and the weight of the aircraft are dependent on its altitude. In this approach, the International Standard Atmosphere model (ISA) [36] is used to determine the ambient density, pressure and temperature. The temperature near tropopause at an altitude of 11 km is modeled using a relation by Kao *et al.* [37], with modifications, to prevent non-differentiability from sudden temperature changes. The variation of the gravitational acceleration of the Earth is calculated via spherical approximation of its shape.

2.2.3 Aerodynamics model

The aerodynamics are modeled as functions of true airspeed V_{TAS} , air density ρ , wing reference area S , with lift F_{A_z} and drag F_{A_x} forces as outputs in the body frame. In this approach, the wind is assumed to be locally parallel to the x -direction of the aircraft in the Earth frame, hence the angle of attack $\alpha = \theta$, the pitch angle in body axes. The force dependence is written as a standard function of the aerodynamic coefficients,

$$\vec{F}_A = \frac{1}{2}\rho V_{\text{TAS}}^2 S \vec{C}_A,\tag{5}$$

where the components C_{A_z} and C_{A_x} denote the lift and drag coefficients, respectively. The methods of computing aerodynamic coefficients used in MARILib¹, an open source multidisciplinary aircraft design library, are utilized to compute the lift and drag coefficients in this study. In particular, the lift coefficient is modeled as a linear function of the angle of attack and the lift curve slope is determined by the Polhamus formula [38] within applicable bounds, with adjustments for high-lift device configurations, using the method presented in Dubs [39]. The parameters of the lift coefficient model are the wing aspect ratio, fuselage width, sweepback angle, wing span, and Mach number. The drag coefficient is determined by an empirical drag build-up method of profile drag, induced drag, and wave drag [40]. The main component of profile drag is the friction/form drag considering turbulent flow with compressibility effects for the fuselage, wing, horizontal and vertical stabilizers, and the nacelle [41]. The friction/form drag of miscellaneous devices such as antennas, sensors and tapered fuselage tail cone drag are also considered. The induced drag is calculated

¹<https://github.com/marilib/MARILib>

considering the span efficiency factor with fuselage interference corrections [42]. The wave drag is derived based on Korn’s relation [43]. The parameters of the drag coefficient model are the lift coefficient, wing reference area, wing wetted area, wing span, mean chord length, fuselage width, fuselage tail cone length, nacelle length, wing aspect ratio, and Mach number.

2.2.4 Propulsion model

In this work, we use a propulsion model that computes engine thrust F_T and specific fuel consumption (SFC, c_T) using the Mach number M , bypass ratio of the engine BPR , throttle factor τ , reference thrust $F_{T_{\text{ref}}}$, and altitude h as inputs. The engine thrust is determined by the empirical formula used in MARILib,

$$F_T = \tau F_{T_{\text{ref}}} \sigma(h)^{0.75} \left[0.475M^2 + 0.091 \left(\frac{BPR}{10} \right)^2 - 0.283M \times \frac{BPR}{10} - 0.633M - 0.081 \frac{BPR}{10} + 1.192 \right], \quad (6)$$

where σ is the air density ratio relative to the sea level value, which is a function of altitude.

The specific fuel consumption is computed using a surrogate model built from an engine data-set available via the open-source library SMT: Surrogate Modeling Toolbox [44]. Specifically for this study, we use engine data pertaining to Boeing 777 aircraft.² The surrogate modeling technique used is the regularized minimal-energy tensor product cubic-splines (RMTTC) [45] model for its fast prediction time with large training data-sets and high accuracy. The model predicts SFC with given throttle factor, Mach number, and altitude. Fig. 6 presents SFC contours with respect to altitude and Mach number at $\tau = 0.56$, 0.688 , which correspond to cruise and climb throttle settings, respectively. The use of this surrogate also reduces computational time compared to a fully physics-based engine cycle analysis, while providing higher accuracy compared to empirical formulas.

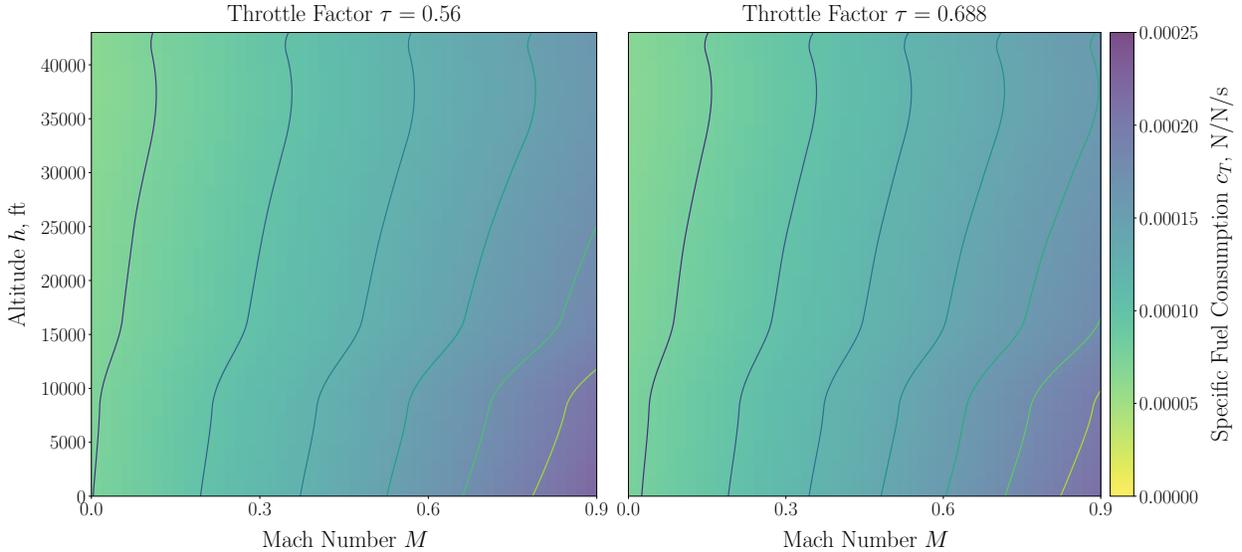


Figure 6: Surrogate model contours for the engine SFC.

²https://smt.readthedocs.io/en/latest/_src_docs/examples/b777_engine/b777_engine.html

2.2.5 Mass model

The mass model computes the aircraft mass for fuel consumption estimation, and the weight for solution of the force equilibrium. The aircraft mass m at the current time-step t_n is obtained by subtracting the mass of fuel consumed from the aircraft mass at the previous time-step t_{n-1} , using the current engine thrust and SFC. The mass is then multiplied by the current gravitational acceleration as shown in Eq. (7) to obtain the weight W_n for evaluation in the force equilibrium equation Eq. (1),

$$W_n = m_n \vec{g}(h_n) = \left[m_{n-1} - \frac{F_T(t_n) \times c_T(t_n)}{g(t_n)} (t_n - t_{n-1}) \right] \vec{g}(h_n), \quad (7)$$

where $n = 1, \dots, K$. Hence the total fuel consumption m_{fuel} , defined as the mass of fuel consumed throughout the entire flight, is the mass at takeoff m_0 subtracted by the mass at the end of the simulation m_{LW} , i.e., the landing mass corresponding to time t_K , as shown in Eq. (8),

$$m_{\text{fuel}} = m_0 - m_{\text{LW}} = \int_0^{t_K} \frac{F_T(t) \times c_T(t)}{g(t)} dt. \quad (8)$$

2.3 Flight segments and constraint formulations

The flight segments defined by the segmentation model described in Section 2.1 are used as the boundary conditions and constraints for the integration in the flight dynamics model presented in Section 2.2.1. In this procedure, the terminal condition of the previous segment is set as the initial condition for the subsequent segment to ensure a physically valid flight mission profile.

The constraints for each flight segment are detailed subsequently to determine the angle of attack α , throttle factor τ , and the high-lift device configuration η at the various discretized points in the different segments. The high-lift device configuration corresponds to the change in the lift-curve slope with deployment of flaps and slats. Two types of problems are formulated based on the flight conditions of the segments, namely the constant speed or Mach conditions by nonlinear solutions as described in Section 2.3.1, and accelerated conditions by optimizations as described in Section 2.3.2.

The XDMS presented in Fig. 7 depicts the nonlinear constraint and optimization formulations used in the flight simulation as an extension of Fig. 5 augmented by the data model. Depending on the flight phase, the corresponding constraints are imposed on the simulation and evaluated. Specifically, the relevant inputs to the flight dynamics, aerodynamics, and propulsion model blocks are set by the corresponding segment constraints in the form of residual equations using the data model. The solutions and optimizations of equations are performed using the SciPy package³.

Fig. 8 shows the vertical speed distribution, based on flight data, of each flight segment classified using the segmentation model described in Section 2.1. The black bars shown in the accelerated climb, decelerated descent, constant CAS climb and descent, and constant Mach climb and descent segment graphs are representative of the average vertical speed for each segment. The bin with the highest frequency is the mode of the distribution. We can observe that the average vertical speed almost coincides with the corresponding mode value for accelerated climb, decelerated descent, constant CAS climb and descent, and constant Mach climb and descent segments. As an example, the representative vertical speeds of the takeoff, constant Mach cruise, and landing segments can be set to zero, where the highest frequencies of occurrences are observed. This observation suggests that the average vertical speed of a segment is also the most frequently operated vertical speed, and hence can be used as the representative vertical speed for that particular segment. In this study,

³<http://www.scipy.org/>

the flight path of the aircraft is constrained by setting the representative average vertical speed as the target value for the residual equation of each segment presented subsequently.

2.3.1 Constant speed or Mach conditions

For unaccelerated segments, the residual equations of the form presented in Eq. (9) are solved for the angle of attack α and throttle factor τ ,

$$\mathcal{R}_1 \left(\begin{bmatrix} f_1 - f_1^* \\ f_2 - f_2^* \end{bmatrix}; \alpha, \eta, \tau \right) = 0, \quad (9)$$

where f_1 and f_2 are quantities of interest computed from the flight simulation. These quantities are certain state variables corresponding to a given segment by solving the discretized form of Eq. (1) for each time-step in the segment. The suitable f_1 and f_2 for each segment will be described below. The asterisks denote the target values, which are determined from the data processing. We set $\eta = 0$ as data reveal that pilots do not typically use high-lift devices in these segments. We assume that the equivalent airspeed is equal to the calibrated airspeed V_{CAS} . The calculation setup for segments belonging to this category is described briefly below. Similar climb and descent segments (e.g. constant Mach or constant CAS) are described together for conciseness.

Constant CAS climb and descent: In these phases, V_{CAS} remains constant as the altitude increases. Consequently, V_{TAS} and M increase, following the relation $V_{CAS} = V_{TAS} \times \sigma(h) = M \times \beta(h) \times \sigma(h)$, where the air density ratio σ and speed of sound β decrease as the altitude increases in the troposphere, hence M increases, and vice versa for the descent segment. As such, the targets are set as $f_1 = V_{CAS}$, $f_2 = V_z$. The termination criteria for the climb and descent segments are set to a Mach number and altitude requirement respectively corresponding to the initial value of the subsequent segment.

Constant Mach climb and descent: In these phases, the Mach number remains constant as the altitude increases. As the speed of sound decreases with increasing altitude, the true airspeed decreases in order to maintain the constant Mach number during climb, and vice versa for the descent segment. Hence the targets are set as $f_1 = M$, $f_2 = V_z$. The termination criteria for both climb and descent are the target altitude h^* corresponding to the initial value of the next segment.

Constant Mach cruise: In this phase, the Mach number and altitude remain constant, hence $f_1 = M$, $f_2 = h$. The termination condition for this segment is the flight time satisfying the range requirements explained in Section 2.3.3.

2.3.2 Accelerated conditions

In the following conditions, the optimization problem presented in Table 2 is solved for the angle of attack, high-lift device configuration, and throttle factor, where f_1 and f_2 are the quantities of interest computed from the flight simulation corresponding to a given segment, and those denoted with asterisks are the corresponding values determined from data processing. The high-lift device configuration value is explicitly varied for these segments, as analyses from data indicate use of high-lift devices.

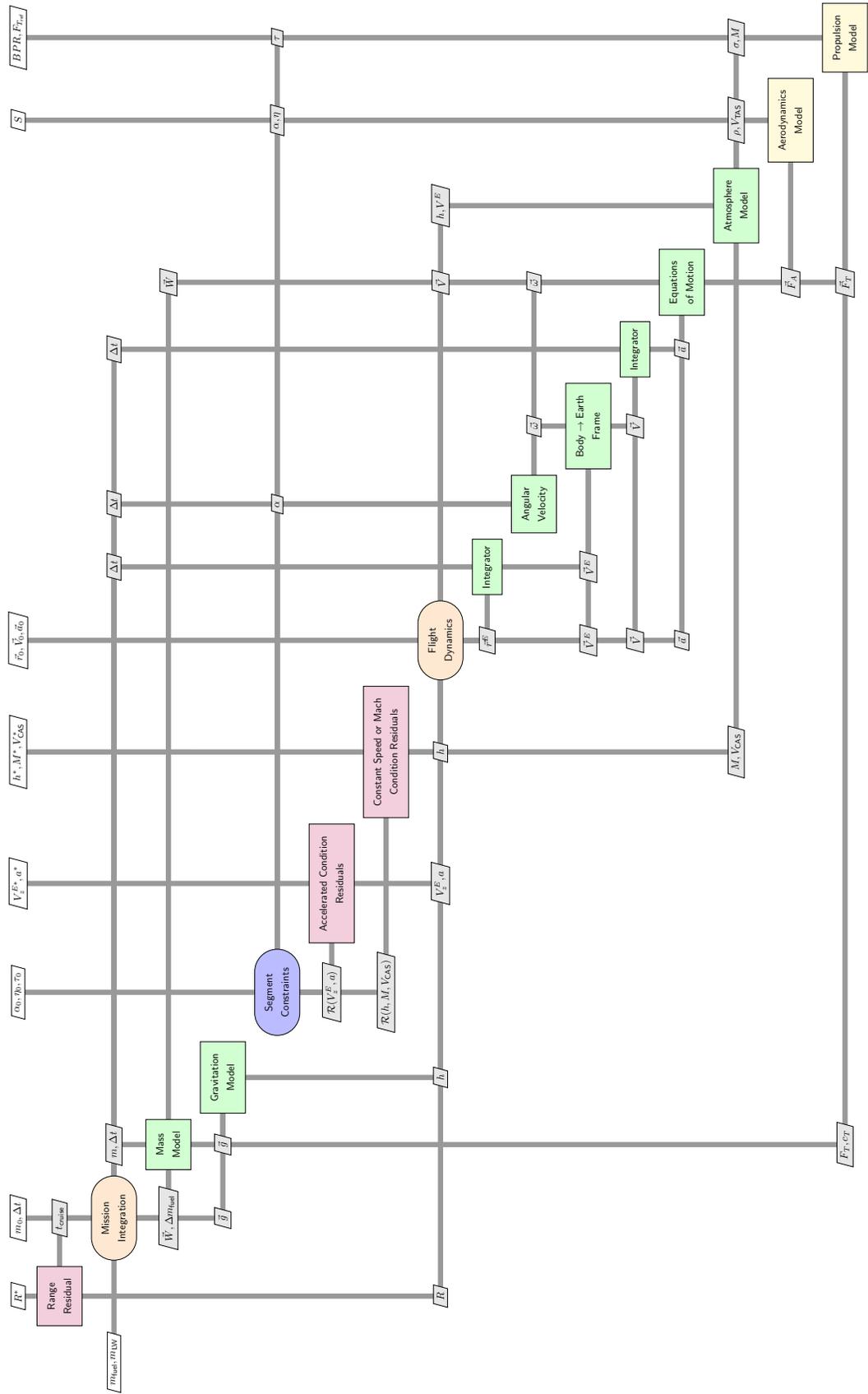


Figure 7: XDSM for the flight simulation methodology.

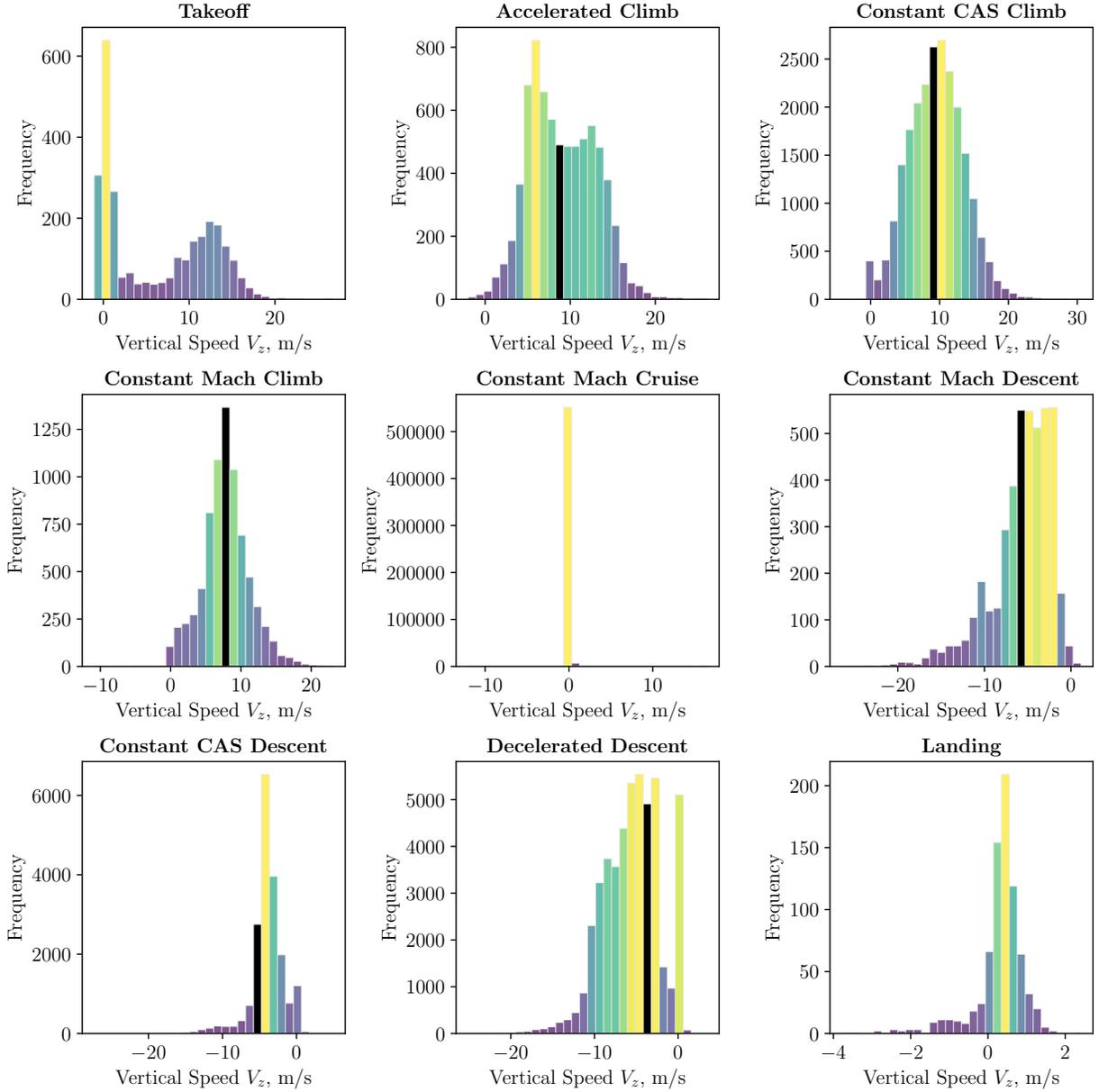


Figure 8: Vertical speed distributions for the various segments.

Table 2: Optimization problem formulation for accelerated flight condition segments.

Optimization	Function variables	Description
Design variables	Minimize $ f_1 - f_1^* $	Residual equation for objective
	α	Angle of attack
	η	High-lift device configuration
Constraints	τ	Throttle factor
	Residual equation for constraint $ f_2 - f_2^* = 0$	Residual equation for constraint
	$\alpha_{\min} \leq \alpha \leq \alpha_{\max}$	Angle of attack range
	$\eta_{\min} \leq \eta \leq \eta_{\max}$	High-lift device configuration values
	$\tau_{\min} \leq \tau \leq \tau_{\max}$	Throttle factor bounds for segment

Accelerated climb and decelerated descent: The acceleration of the aircraft is maintained at a constant non-zero value in these phases. The segment flight time is calculated using the vertical speeds and altitudes at the initial and final points of the segment. The difference in true airspeeds between the initial and final points of the segment is divided by the flight time to obtain the target acceleration. The climb and descent segments terminate when they reach or exceed the target true airspeed, which then becomes the initial condition of the next segment, hence $f_1 = V_{TAS}$. As the final altitude is used to compute the target acceleration a , the altitude is implicitly considered in the termination criteria, hence $f_2 = h$. The termination criterion for the descent segment *specifically before landing* is expressed in terms of the stall speed V_{stall} , when the speed decreases to the approach speed given by $V_{TAS}^* = 1.3V_{stall}$ [7].

Takeoff: The design variables for takeoff are determined for the initial point by computing the simulations using the optimization schema for fixed values until the termination criteria are satisfied. This algorithm is then repeated for the subsequent points. The termination criteria are the true airspeed, and implicitly the altitude, hence $f_1 = V_{TAS}$, $f_2 = h$.

Landing: The algorithm for the landing segment is the same as the one for takeoff, except for the repetition on the subsequent points. The termination criteria are the touch-down speed and final altitude, hence $f_1 = V_{TAS}$, $f_2 = h_2$. The touch-down speed is calculated as $V_{TAS}^* = 1.15V_{stall}$ [7].

2.3.3 Cruise range determination

The range of the cruise segment is an unknown value that must be determined by performing cruise, descent and landing simulations to see if the total range requirement for the mission is satisfied. To address this, we solve the following residual equation for the cruise flight time t_{cruise} , which determines the range of the dominant cruise segment,

$$\mathcal{R}_2\left(\frac{R - R^*}{R^*}; t_{cruise}\right) = 0, \quad (10)$$

where R^* is a target flight range and R is the computed mission range covered upon completion of all the segments, in terms of ground distance. The nonlinear equation is solved using a bisection method, in which the appropriate cruise flight time is determined by calculating whether the simulations over- or under-shoot the range requirements. Specifically, this strategy evaluates the constant Mach cruise segment and any subsequent segments until landing, which solves Eq. (9) correspondingly for each segment. R is computed after landing, and the residual \mathcal{R}_2 is then evaluated against the target R^* . The process is repeated until convergence is obtained to the specified tolerance, and hence the necessary flight time for the cruise segment t_{cruise} is determined. The flight simulation is completed when this process of cruise range determination is finished, and the results are then evaluated for performance analysis of the aircraft.

3 Results and discussion

This section provides the results and performance analyses of flight simulations based on the proposed methodology for three different routes, i.e., short-haul flight (Section 3.1), medium-haul flight (Section 3.2), and long-haul flight (Section 3.3). The origin of all the routes is set as Hong Kong International Airport (HKG). The respective destinations are Taiwan Taoyuan International Airport (TPE) for short-haul, Singapore Changi Airport (SIN) for medium-haul, and Heathrow Airport, London (LHR) for long-haul. The parameters of an aircraft used for the simulations are shown in Table 3, with specifications similar to a Boeing 777-300ER with two GE90-115B engines.

Table 3: Aircraft parameters

Parameter	Value
Reference Aircraft	Boeing 777-300ER
Propulsion Type	Turbofan
Bypass Ratio	9
Number of Engines	2
Number of Passengers	396
Wing Area	436.8 m ²

3.1 Short-haul flight

The mission specification for the HKG-TPE flight sector is shown in Table 4, determined from the segmentation model. The table indicates the boundary conditions imposed on altitudes h_i and speeds V_i , indexed at the endpoints of N segments in sequence for $i = 0, \dots, N$. The speeds are expressed in different measures of true airspeed in knots (KTAS), such as CAS and Mach number, corresponding to values determined from data processing for the appropriate segments.

In this tabular presentation, explicitly assigned variables indicate which boundary values are set using flight data. Unassigned variables are either inherited from the corresponding final values of the previous segment or via the imposition of boundary conditions on other variables, which are calculated during the simulation. Variables with asterisks indicate target values for final boundary conditions of a segment. These targets may not be exactly satisfied due to numerical errors in the solution of the residual equations during the integration, hence the tolerance for the absolute error is set to 10^{-3} to obtain sufficient accuracy with low computational time. For illustration purposes, consider the constant Mach climb phase in Table 4. The initial values h_3, V_3 of this segment are inherited from the previous values determined in the constant CAS climb phase. The altitude increases to the explicitly assigned target h_4^* while the corresponding target speed V_4 is unassigned. This target speed is calculated via solution of the residual equations for this phase, which constrains the Mach number to be constant. As the altitude increases, the density decreases, which changes the true airspeed value from V_3 to V_4 , while maintaining the same Mach number.

Fig. 9 depicts the results of the simulation with the design variables colored in red, the state variables colored in green, and other quantities of interest such as the propulsion components colored in yellow. The variations of the design variables α, η, τ are physically reasonable and indicate that the aerodynamics and propulsion models are sufficiently applicable for analyses. They also indicate that the procedure for solving the residual equations determines the appropriate values of the design variables. During cruise, the phase with the highest altitude in the figure, the values of angle of attack and throttle factor computed from the simulation are overestimated with reference to data, which provide average values of approximately 2.5° and 0.5 respectively. This discrepancy

Table 4: Short-haul (HKG-TPE) flight mission specification. (Note that the takeoff mass is masked for confidentiality reasons, as stipulated in the agreement with our airline partner.)

Segment	Altitudes, ft		Speeds	
	<i>Initial</i>	<i>Final</i>	<i>Initial</i>	<i>Final</i>
Takeoff	$h_0 = 0$	$h_1^* = 1\,741$	$V_0 = 97$ KTAS	$V_1^* = 210$ KTAS
Accelerated Climb	h_1	$h_2^* = 13\,305$	V_1	$V_2^* = 373$ KTAS
Constant CAS Climb	h_2	h_3	V_2	$V_3^* = \text{Mach } 0.85$
Constant Mach Climb	h_3	$h_4^* = 36\,962$	V_3	V_4
Constant Mach Cruise	h_4	h_4	$V_4 = \text{Mach } 0.84$	V_4
Constant Mach Descent	$h_5 = h_4$	$h_6^* = 35\,722$	$V_5 = \text{Mach } 0.85$	V_6
Decelerated Descent	h_6	$h_7^* = 33\,602$	V_6	$V_7^* = 471$ KTAS
Constant CAS Descent	h_7	$h_8^* = 22\,078$	V_7	V_8
Decelerated Descent	h_8	$h_9^* = 21\,495$	V_8	$V_9^* = 357$ KTAS
Constant CAS Descent	h_9	$h_{10}^* = 13\,683$	V_9	V_{10}
Decelerated Descent	h_{10}	$h_{11}^* = 10\,827$	V_{10}	$V_{11}^* = 266$ KTAS
Constant CAS Descent	h_{11}	$h_{12}^* = 6\,058$	V_{11}	V_{12}
Decelerated Descent	h_{12}	$h_{13}^* = 31$	V_{12}	$V_{13}^* = 204$ KTAS
Landing	h_{13}	$h_{14}^* = 0$	V_{13}	$V_{14}^* = 0$
Total Flight Range, R^*				539 nm
Takeoff Mass, m_0				4 ████████ lbs

is attributed to the low level of fidelity in the aerodynamics and propulsion models, or difference in throttle setting. However, the accuracy of the individual models does not significantly affect the estimation of other state variables, which are not computed from residual equations, and the accuracy of fuel prediction in particular. This is because the entire flight profile is indirectly constrained by solution of the residual equations, which affects the force balance of the aircraft during flight, thus contributing to the accuracy of fuel prediction. This characteristic is reflected in the computed trajectory, colored in blue, when compared to the actual trajectory of the aircraft from QAR data, colored in red.

A transient jump in the state variables is observed between transition from the constant Mach climb to the constant Mach cruise segment at approximately 12000 seconds. This is due to limitations in matching certain boundary conditions of the model, which treats the changes in the state variables between these segments as instantaneous. Similar transients are also observed in the accelerated climb and decelerated descent segments due to the inner optimization simultaneously adjusting the throttle factor, angle of attack, and high-lift device configuration values; these may indicate that the solutions to the residuals are non-unique under these conditions. Particularly, the transient is caused by the jump in Mach numbers between the final and initial values V_4 of the corresponding segments. The performance analysis presented in Section 3.4 indicates that any errors due to these transients are negligible.

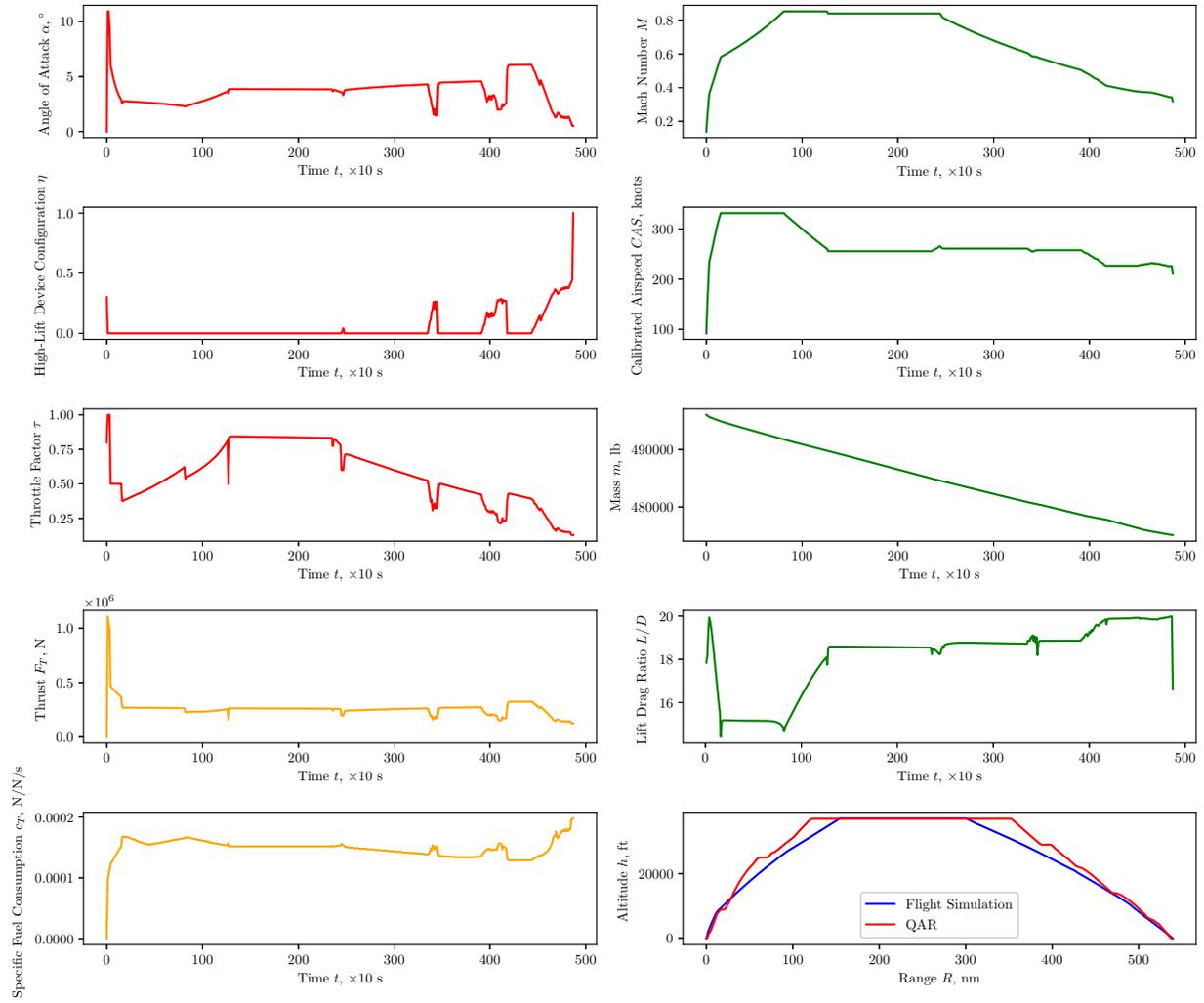


Figure 9: Simulation results for the short-haul sector (HKG-TPE).

3.2 Medium-haul flight

The mission specification for the HKG-SIN flight sector is shown in Table 5, and the results of the simulation are depicted in Fig. 10, following the same format as the previous sector’s presentation. The computed trajectory graphically appears to replicate the actual trajectory from QAR data more closely than the HKG-TPE results; however, this is only due to the scale of the cruise segment being much longer comparatively.

Table 5: Medium-haul (HKG-SIN) flight mission specification. (The takeoff mass is masked for confidentiality reasons.)

Segment	Altitudes, ft		Speeds	
	<i>Initial</i>	<i>Final</i>	<i>Initial</i>	<i>Final</i>
Takeoff	$h_0 = 0$	$h_1^* = 1\,723$	$V_0 = 97$ KTAS	$V_1^* = 216$ KTAS
Accelerated Climb	h_1	$h_2^* = 6\,324$	V_1	$V_2^* = 336$ KTAS
Constant CAS Climb	h_2	h_3	V_2	$V_3^* = \text{Mach } 0.85$
Constant Mach Climb	h_3	$h_4^* = 36\,016$	V_3	V_4
Constant Mach Cruise	h_4	h_4	$V_4 = \text{Mach } 0.85$	V_4
Constant Mach Descent	$h_5 = h_4$	$h_6^* = 32\,997$	$V_5 = \text{Mach } 0.84$	V_6
Constant Mach Cruise	h_6	h_6	$V_6 = \text{Mach } 0.83$	V_6
Constant CAS Descent	$h_7 = h_6$	$h_8^* = 32\,056$	$V_7 = V_6$	V_8
Constant Mach Cruise	h_8	h_8	$V_8 = \text{Mach } 0.80$	V_8
Constant CAS Descent	$h_9 = h_8$	$h_{10}^* = 21\,782$	$V_9 = V_8$	V_{10}
Decelerated Descent	h_{10}	$h_{11}^* = 20\,485$	V_{10}	$V_{11}^* = 349$ KTAS
Constant CAS Descent	h_{11}	$h_{12}^* = 18\,045$	V_{11}	V_{12}
Decelerated Descent	h_{12}	$h_{13}^* = 10\,827$	V_{12}	$V_{13}^* = 300$ KTAS
Constant CAS Descent	h_{13}	$h_{14}^* = 7\,619$	V_{13}	V_{14}
Decelerated Descent	h_{14}	$h_{15}^* = 15$	V_{14}	$V_{15}^* = 210$ KTAS
Landing	h_{15}	$h_{16}^* = 0$	V_{15}	$V_{16}^* = 0$
Total Flight Range, R^*				1 487 nm
Takeoff Mass, m_0				5 [REDACTED] lbs

In the figure, it is interesting to note that the angle of attack inversely correlates with the value of the high-lift device configuration for segments in which the latter is not fixed. This trend sustains the required lift with the decreasing throttle value to satisfy the acceleration constraints as the weight of the aircraft decreases with fuel burn. With such detailed analysis of flight performance factors, the effects on operational variables, which are inputs of the flight simulation, can hence be assessed and optimized in flight time or fuel consumption aspects.

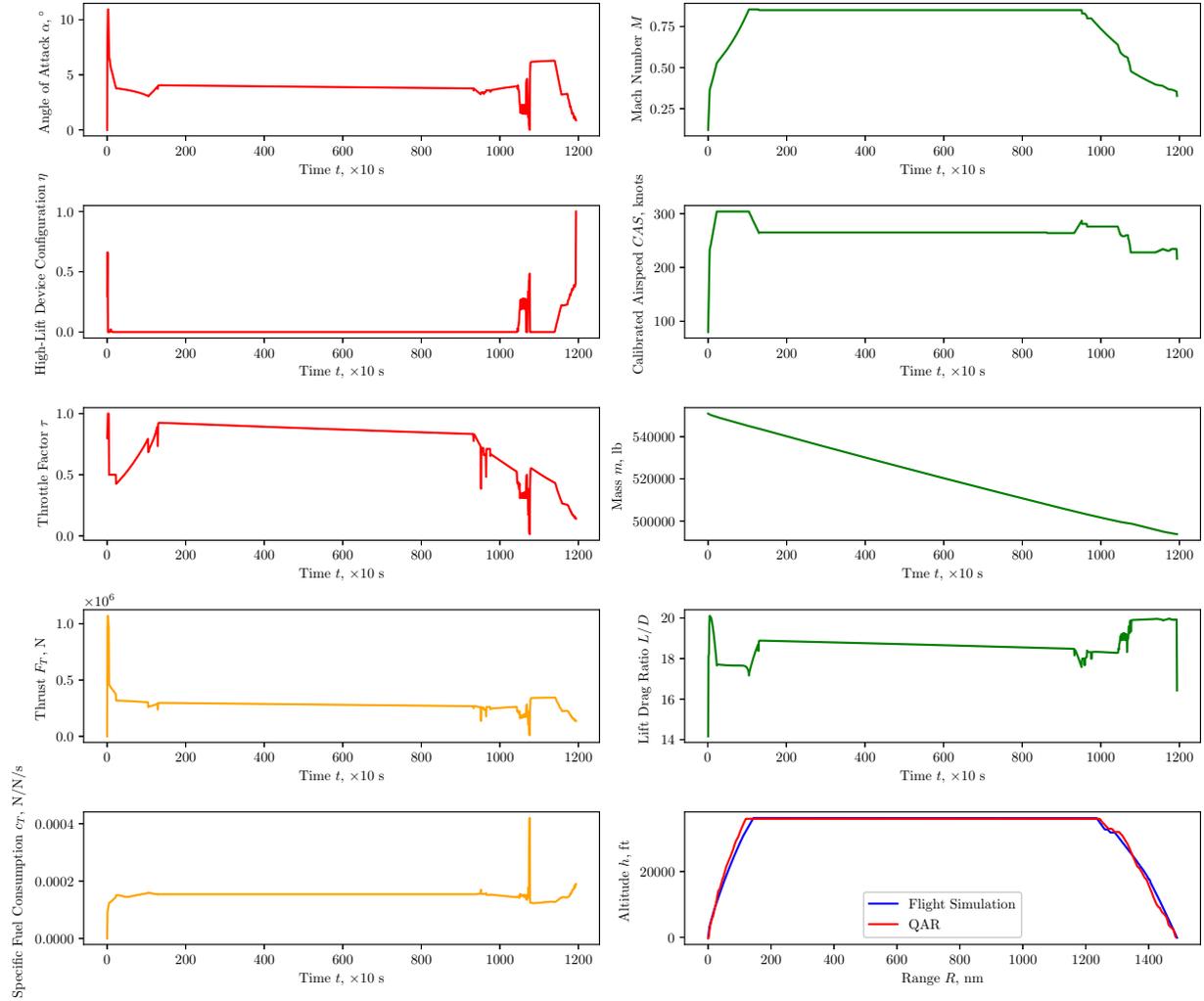


Figure 10: Simulation results for the medium-haul sector (HKG-SIN).

3.3 Long-haul flight

The mission specification for the HKG-LHR flight sector is shown in Table 6, and the results of the simulation are depicted in Fig. 11, following the same format. Unlike the short-haul and medium-haul flights, in which the range residual computation only affects the length of one cruise segment, the main cruise section of the long-haul flight is divided into three cruise segments with three constant Mach climb segments replicating step climbs. As the exact location and time of the step climbs are not important in this study, the duration of each of these cruise segments is assumed to be the same as the others' to utilize the cruise range determination method presented in Section 2.3.3.

Under these considerations, the computed trajectory of the aircraft matches the actual trajectory from QAR data in the cruise and step climb segments to an acceptable degree. These performance factors shed important physical insight into real-world aircraft operations, which are not typically available when using simplified methods that do not model segment-by-segment operations. This information is useful in determining the flight operating envelope in aircraft design optimization, though the optimization is beyond the scope of this paper.

Table 6: Long-haul (HKG-LHR) flight mission specification. (The takeoff mass is masked for confidentiality reasons.)

Segment	Altitudes, ft		Speeds	
	<i>Initial</i>	<i>Final</i>	<i>Initial</i>	<i>Final</i>
Takeoff	$h_0 = 0$	$h_1^* = 2\,913$	$V_0 = 97$ KTAS	$V_1^* = 237$ KTAS
Accelerated Climb	h_1	$h_2^* = 16\,788$	V_1	$V_2^* = 418$ KTAS
Constant CAS Climb	h_2	h_3	V_2	$V_3^* = \text{Mach } 0.77$
Constant Mach Climb	h_3	h_4	V_3	V_3
Constant CAS Climb	h_4	h_5	$V_4 = V_3$	$V_5^* = \text{Mach } 0.84$
Constant Mach Climb	h_5	h_5	V_5	V_5
Constant Mach Climb	$h_6 = h_5$	$h_7^* = 32\,110$	$V_6 = V_5$	V_7
Constant Mach Climb	h_7	h_7	$V_7 = \text{Mach } 0.82$	V_7
Constant Mach Climb	$h_8 = h_7$	$h_9^* = 33\,980$	$V_8 = \text{Mach } 0.81$	V_9
Constant Mach Climb	h_9	h_9	$V_9 = \text{Mach } 0.81$	V_9
Constant Mach Climb	$h_{10} = h_9$	$h_{11}^* = 35\,998$	$V_{10} = \text{Mach } 0.81$	V_{11}
Constant Mach Climb	h_{11}	h_{11}	$V_{11} = \text{Mach } 0.81$	V_{11}
Constant CAS Descent	$h_{12} = h_{11}$	$h_{13}^* = 34\,051$	$V_{12} = V_{11}$	V_{13}
Decelerated Descent	h_{13}	$h_{14}^* = 31\,284$	V_{13}	$V_{14}^* = 399$ KTAS
Constant CAS Descent	h_{14}	$h_{15}^* = 25\,163$	V_{14}	$V_{15}^* = \text{Mach } 0.57$
Constant Mach Descent	h_{15}	$h_{16}^* = 23\,259$	V_{15}	V_{16}
Constant CAS Descent	h_{16}	$h_{17}^* = 7\,990$	V_{16}	V_{17}
Decelerated Descent	h_{17}	$h_{18}^* = 727$	V_{17}	$V_{18}^* = 213$ KTAS
Landing	h_{18}	$h_{19}^* = 0$	V_{18}	$V_{19}^* = 0$
Total Flight Range, R^*				5 351 nm
Takeoff Mass, m_0				7 ████████ lbs

3.4 Fuel consumption and flight time analyses and comparisons

The Bréguet range and endurance equations shown in Eq. (11) are normally used to compute fuel consumption and flight time in the cruise phase with initial weight W_i and final weight W_f . These equations are usually considered first in preliminary aircraft design before adding more complex phases such as accelerated conditions.

$$\begin{aligned}
 m_{\text{fuel, Bréguet}} &= m_0 \left(1 - \exp \left[- \frac{c_T \times R}{V_{\text{TAS, cruise}} \times (L/D)_{\text{cruise}}} \right] \right) \\
 t_{\text{Bréguet}} &= \frac{1}{c_T} \left(\frac{L}{D} \right)_{\text{cruise}} \ln \left(\frac{W_i}{W_f} \right)
 \end{aligned} \tag{11}$$

Fig. 12 shows the fuel weight fractions predicted from the flight simulation, the Bréguet range equation, and OpenAP⁴, which is an open-source aircraft performance model. These predictions are compared against QAR data for reference. The accuracy of the fuel consumption computed from the flight simulations is greater than those of the Bréguet range equation and OpenAP for all the missions considered in comparison to the data. The Bréguet range equation underestimates the fuel consumed, while the simulations overestimate the value. The latter estimation is more desirable as a factor of safety. The accuracy of the flight time computed from the flight simulations

⁴<https://github.com/TUDeIft-CNS-ATM/openap>

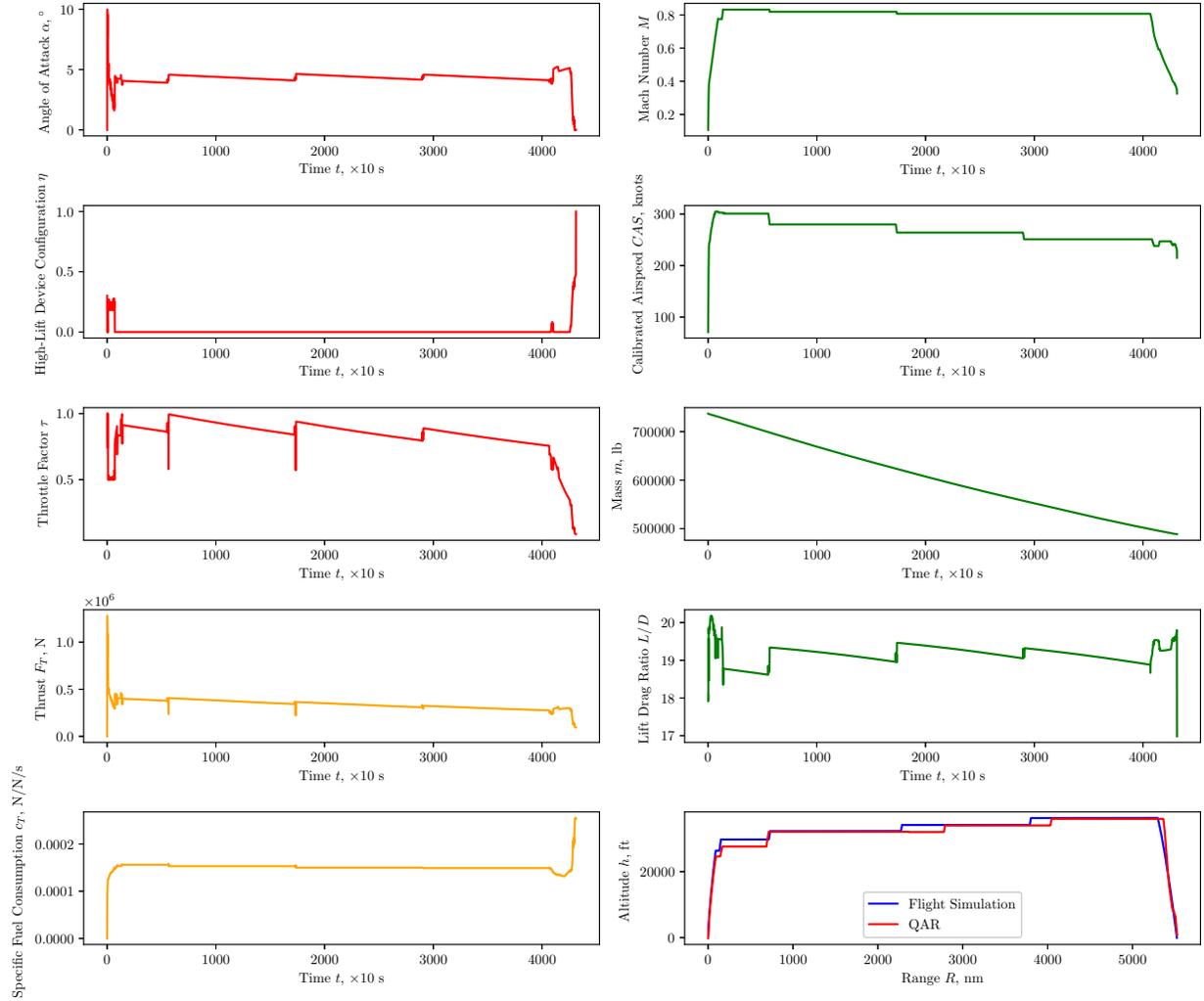


Figure 11: Simulation results for the long-haul sector (HKG-LHR).

is also similarly greater, which is attributed to the residual formulation presented in Section 2.3.3. The only exception is the computed error for the long-haul flight time, which matches the error computed from OpenAP’s result. This is expected due to the rotation of the Earth not being modeled, with flights of longer range resulting in larger errors.

Fig. 12 also demonstrates that the developed flight simulation model is notably superior to other prediction models in predicting fuel consumption when the aircraft operates further away from its design mission. Accurate fuel prediction is important for airlines, for the purposes of fuel budgeting and fuel economy assessment. The Boeing 777-300ER is designed primarily for long-haul operations, and among the three flight sectors considered in this study, the HKG-LHR flight sector is the one closest to the intended design mission. While the performances of the three prediction models are comparable for this flight sector, the fuel prediction error of our proposed flight simulation model is notably smaller than the other models’ for the other two sectors, whose flight ranges are shorter. These results further emphasize the importance of considering actual flight trajectories and operation parameters for realistic aircraft performance evaluations.

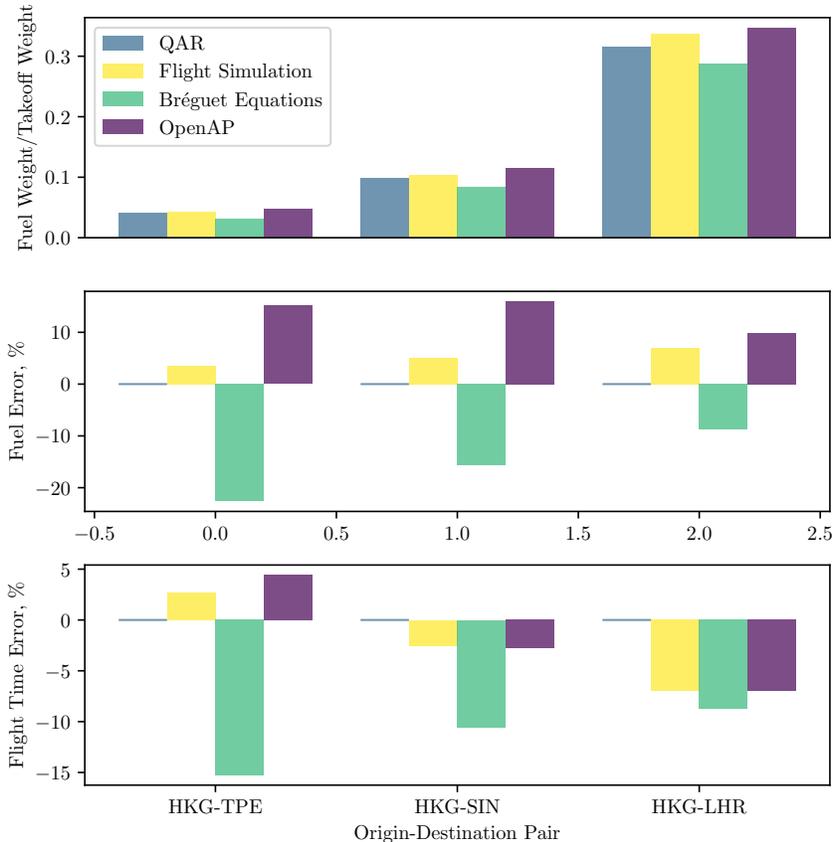


Figure 12: Fuel consumption and flight time comparisons between QAR data, our proposed flight simulation methodology, the Bréguet range and endurance equations, and the OpenAP model.

4 Conclusion

In this paper, we present a data-enhanced flight simulation physics model to evaluate aircraft performance with reduced information about the aircraft characteristics. The methodology is developed for evaluating aircraft performance parameters via the formulation of dynamic flight simulations, which are enhanced by use of aircraft operational data from real-world flights across various sectors. The flight data are clustered into segments by XGBoost classification, which are then analyzed to develop mission specifications. Further analyses of the data are also performed to determine the appropriate boundary conditions and constraints in the differential equations of motion governing the flight dynamics. The formulation is simple enough to avoid the requirement of a control systems architecture, and yet complex enough to allow modular development with aerodynamics and propulsion models of varying fidelity. The flight performance analysis of an aircraft similar to a Boeing 777-300ER with GE90-115B engine specifications is conducted using this methodology in various operating scenarios, and the results are validated against QAR data. The evaluations of fuel burn and flight time, which are the main performance parameters considered in this study, yield results with lesser errors against these reference data, as compared to results from the Bréguet range and endurance equations, and the OpenAP flight performance model. The results and observations presented in this paper demonstrate the benefits of incorporating data processing into a physics-based simulation model. The operational performance captured by data helps impose realistic constraints on the model formulation, hence ensuring a valid simulation of the physical

system even when some detailed information is absent, while maintaining computational efficiency. Such a hybrid approach, which combines the objectivity of data and interpretability of physics-based models, can lend itself to model other data-intensive engineering problems outside aircraft performance analysis, such as weather modeling. The developed methodology will also be suitable for incorporating actual aircraft operations in preliminary aircraft design frameworks, establishing a systematic link between aircraft design and air transportation operations. The flight conditions generated from completing the flight performance analyses, for instance, can be used to determine the relevant flight-operation regime at which the aircraft performance is optimized. Such a formulation does not rely on nominal assumptions, and therefore can yield a more realistic aircraft design. This data-driven aircraft design optimization problem formulation is a subject of future work.

Declaration of competing interest

The authors declare no competing interests in the development of this research.

Acknowledgment

The authors would like to thank Cathay Pacific Airways Limited for providing the data used in this study under the Data Partnership Agreement between the airline and the Department of Mechanical and Aerospace Engineering, HKUST. The work is supported by the HKUST Start-Up Grant (Project No. R9354). The authors additionally thank the anonymous reviewers of this paper for their helpful suggestions.

References

- [1] M. Husemann, K. Schäfer, E. Stumpf, Flexibility within flight operations as an evaluation criterion for preliminary aircraft design, *Journal of Air Transport Management* 71 (2018) 201–214. doi:10.1016/j.jairtraman.2018.04.007.
- [2] L. Garrow, V. Lurkin, How COVID-19 is impacting and reshaping the airline industry, *Journal of Revenue and Pricing Management* 20 (1) (2021) 3–9. doi:10.1057/s41272-020-00271-1.
- [3] The International Civil Aviation Organization, Economic impacts of COVID-19 on civil aviation, Available at <https://www.icao.int/sustainability/Pages/Economic-Impacts-of-COVID-19.aspx>.
- [4] X. Sun, S. Wandelt, A. Zhang, How did COVID-19 impact air transportation? A first peek through the lens of complex networks, *Journal of Air Transport Management* 89 (2020) 101928. doi:10.1016/j.jairtraman.2020.101928.
- [5] A. T. Bureau, Effects of novel coronavirus (COVID-19) on Civil Aviation: Economic Impact Analysis, Available at https://www.icao.int/sustainability/Documents/COVID-19/ICAO_Coronavirus_Econ_Impact.pdf (2020).
- [6] A. D. Marco, V. Trifari, F. Nicolosi, M. Ruocco, A simulation-based performance analysis tool for aircraft design workflows, *Aerospace* 7 (11) (2020) 155. doi:https://doi.org/10.3390/aerospace7110155.
- [7] D. Raymer, *Aircraft Design: A Conceptual Approach 5e and RDSWin STUDENT*, American Institute of Aeronautics and Astronautics, Inc., 2012.

- [8] S. Ackert, Aircraft payload-range analysis for financiers, Available at https://catsr.vse.gmu.edu/SYST660/aircraft_payload_range_analysis_for_financiers__v1.pdf (2013).
- [9] M. S. Ryerson, M. Hansen, J. Bonn, Fuel consumption and operational performance, in: 9th USA/Europe Air Traffic Management Research and Development Seminar, 2011, pp. 305–314. URL https://www.researchgate.net/profile/Mark-Hansen-7/publication/228519994_Fuel_Consumption_and_Operational_Performance/links/57a1c3ed08ae5f8b258a0397/Fuel-Consumption-and-Operational-Performance.pdf
- [10] J. Yanto, R. P. Liem, Aircraft fuel burn performance study: A data-enhanced modeling approach, *Transportation Research Part D: Transport and Environment* 65 (2018) 574–595. doi:10.1016/j.trd.2018.09.014.
- [11] L. Kang, M. Hansen, Improving airline fuel efficiency via fuel burn prediction and uncertainty estimation, *Transportation Research Part C: Emerging Technologies* 97 (2018) 128–146. doi:10.1016/j.trc.2018.10.002.
- [12] X. Zhu, L. Li, Flight time prediction for fuel loading decisions with a deep learning approach, *Transportation Research Part C: Emerging Technologies* 128 (2021) 103179. doi:10.1016/j.trc.2021.103179.
- [13] G. Wang, K. Liu, H. Chen, Y. Wang, Q. Zhao, A high-precision method of flight arrival time estimation based on XGBoost, in: 2020 IEEE 2nd International Conference on Civil Aviation Safety and Information Technology (ICCASIT, IEEE, 2020, pp. 883–888. doi:10.1109/ICCASIT50869.2020.9368723.
- [14] Z. Guo, B. Yu, M. Hao, W. Wang, Y. Jiang, F. Zong, A novel hybrid method for flight departure delay prediction using Random Forest Regression and Maximal Information Coefficient, *Aerospace Science and Technology* 116 (2021) 106822. doi:10.1016/j.ast.2021.106822.
- [15] J. V. Pérez-Rodríguez, J. M. Pérez-Sánchez, E. Gómez-Déniz, Modelling the asymmetric probabilistic delay of aircraft arrival, *Journal of Air Transport Management* 62 (2017) 90–98. doi:10.1016/j.jairtraman.2017.03.001.
- [16] B. Yu, Z. Guo, S. Asian, H. Wang, G. Chen, Flight delay prediction for commercial air transport: A deep learning approach, *Transportation Research Part E: Logistics and Transportation Review* 125 (2019) 203–221. doi:10.1016/j.tre.2019.03.013.
- [17] Y. Xu, H. Zhang, Z. Liao, L. Yang, A dynamic air traffic model for analyzing relationship patterns of traffic flow parameters in terminal airspace, *Aerospace Science and Technology* 55 (2016) 10–23. doi:10.1016/j.ast.2016.05.010.
- [18] Y. Lin, J.-w. Zhang, H. Liu, Deep learning based short-term air traffic flow prediction considering temporal–spatial correlation, *Aerospace Science and Technology* 93 (2019) 105113. doi:10.1016/j.ast.2019.04.021.
- [19] D. Schmidt, *Modern Flight Dynamics*, McGraw-Hill Higher Education, 2011.
- [20] T. Goetzendorf-Grabowski, D. Mieszalski, E. Marcinkiewicz, Stability analysis using SDSA tool, *Progress in Aerospace Sciences* 47 (8) (2011) 636–646. doi:10.1016/j.paerosci.2011.08.007.

- [21] R. Takase, J. O. Entzinger, S. Suzuki, Pilot-in-the-loop simulation of simple adaptive fault-tolerant controller, *Aerospace Science and Technology* 106 (2020) 106125. doi:/10.1016/j.ast.2020.106125.
- [22] J. Sun, J. M. Hoekstra, J. Ellerbroek, OpenAP: An open-source aircraft performance model for air transportation studies and simulations, *Aerospace* 7 (8) (2020) 104. doi:10.3390/aerospace7080104.
- [23] L. Bréguet, *Calcul du poids de combustible consommé par un avion en vol ascendant*, Gauthier-Villars, 1920.
- [24] T. Y. Druot, M. Belleville, P. Roches, F. Gallard, N. Peteilh, A. Gazaix, A multidisciplinary airplane research integrated library with applications to partial turboelectric propulsion, in: *AIAA Aviation 2019 Forum*, 2019, p. 3243. doi:10.2514/6.2019-3243.
- [25] Y. Yu, H. Yao, Y. Liu, Aircraft dynamics simulation using a novel physics-based learning method, *Aerospace Science and Technology* 87 (2019) 254–264. doi:10.1016/j.ast.2019.02.021.
- [26] K. Seymour, M. Held, G. Georges, K. Boulouchos, Fuel Estimation in Air Transportation: Modeling global fuel consumption for commercial aviation, *Transportation Research Part D: Transport and Environment* 88 (2020) 102528. doi:10.1016/j.trd.2020.102528.
- [27] W. Randle, C. Hall, M. Vera-Morales, Improved range equation based on aircraft flight data, *Journal of Aircraft* 48 (2011) 1291–1298. doi:10.2514/1.C031262.
- [28] J. Mieloszyk, T. Goetzendorf-Grabowski, Introduction of full flight dynamic stability constraints in aircraft multidisciplinary optimization, *Aerospace Science and Technology* 68 (2017) 252–260. doi:10.1016/j.ast.2017.05.024.
- [29] J. Mieloszyk, A. Tarnowski, A. Tomaszewski, T. Goetzendorf-Grabowski, Validation of flight dynamic stability optimization constraints with flight tests, *Aerospace Science and Technology* 106 (2020) 106193. doi:10.1016/j.ast.2020.106193.
- [30] G. B. Chatterji, Fuel burn estimation using real track data, in: *11th AIAA Aviation Technology, Integration, and Operations (ATIO) Conference, including the AIAA Balloon Systems Conference and 19th AIAA Lighter-Than*, 2011, p. 6881. doi:10.2514/6.2011-6881.
- [31] Y. Lyu, R. P. Liem, Flight performance analysis with data-driven mission parameterization: mapping flight operational data to aircraft performance analysis, *Transportation Engineering* 2 (2020) 100035. doi:10.1016/j.treng.2020.100035.
- [32] T. Chen, C. Guestrin, XGBoost: A Scalable Tree Boosting System, in: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16, ACM, New York, NY, USA, 2016*, pp. 785–794. doi:10.1145/2939672.2939785.
- [33] S. M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, in: I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (Eds.), *Advances in Neural Information Processing Systems 30*, Curran Associates, Inc., 2017, pp. 4765–4774.
 URL <http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf>

- [34] B. Bengfort, R. Bilbro, Yellowbrick: Visualizing the scikit-learn model selection process, *Journal of Open Source Software* 4 (35) (2019) 1075. doi:10.21105/joss.01075.
- [35] A. B. Lambe, J. R. R. A. Martins, Extensions to the Design Structure Matrix for the description of multidisciplinary design, analysis, and optimization processes, *Structural and Multidisciplinary Optimization* 46 (2012) 273–284. doi:10.1007/s00158-012-0763-y.
- [36] ISO/TC 20/SC 6, Standard Atmosphere, Standard, International Organization for Standardization, Geneva, Switzerland (1975).
- [37] J. Kao, J. Hwang, J. R. Martins, J. S. Gray, K. T. Moore, A modular adjoint approach to aircraft mission analysis and optimization, in: 56th AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, 2015, p. 0136. doi:10.2514/6.2015-0136.
- [38] E. C. Polhamus, A simple method of estimating the subsonic lift and damping in roll of sweptback wings, Vol. 1862, National Advisory Committee for Aeronautics, 1949.
URL <https://ntrs.nasa.gov/citations/19930082534>
- [39] F. Dubs, *Aerodynamik der reinen Unterschallströmung*, Vol. 1, Springer-Verlag, 2013.
- [40] M. Jahanmiri, Aircraft drag reduction: an overview, Tech. rep., Chalmers University of Technology (2011).
URL <http://publications.lib.chalmers.se/records/fulltext/137214.pdf>
- [41] R. Finck, USAF (United States Air Force) Stability and Control DATCOM (Data Compendium), Tech. rep., McDonnell Aircraft Co., St. Louis, MO (1978).
- [42] M. Niță, D. Scholz, Estimating the Oswald factor from basic aircraft geometrical parameters, Deutsche Gesellschaft für Luft-und Raumfahrt-Lilienthal-Oberth eV, 2012.
URL http://www.fzt.haw-hamburg.de/pers/Scholz/OPerA/OPerA_PUB_DLRK_12-09-10.pdf
- [43] O. Gur, W. H. Mason, J. A. Schetz, Full-configuration drag estimation, *Journal of Aircraft* 47 (4) (2010) 1356–1367. doi:10.2514/1.47557.
- [44] M. A. Bouhlef, J. T. Hwang, N. Bartoli, R. Lafage, J. Morlier, J. R. R. A. Martins, A Python surrogate modeling framework with derivatives, *Advances in Engineering Software* 135 (2019) 102662. doi:10.1016/j.advengsoft.2019.03.005.
- [45] J. T. Hwang, J. R. Martins, A fast-prediction surrogate model for large datasets, *Aerospace Science and Technology* 75 (2018) 74–87. doi:10.1016/j.ast.2017.12.030.